MAS352 Probability & Analysis Reminder Notes

1 Probability

1.1 Sample spaces and random variables

Probability is about carrying out experiments where the outcome of the experiment cannot be fully known in advance. The simplest example is tossing a (fair) coin – we can be confident that the coin is equally likely to land showing heads or tails, but we can't tell in advance which one will show. If we repeat the experiment, we might get a different outcome.

To represent this mathematically, we need two key ingredients. We need:

- A **sample space** Ω, which is a set that contains every possible result of the experiment.

  For example, if we roll a (six sided) dice, we might take $\Omega = \{1, 2, 3, 4, 5, 6\}$, with one number for each possible outcome of the dice roll.

- A **probability measure** $\mathbb{P}$, which is able to tell us how likely each of the different outcomes inside Ω is.

  In our coin toss, we'd probably take $\Omega = \{H, T\}$, where we use H to represent heads and T to represent tails. Then, $\mathbb{P}[H] = \mathbb{P}[T] = \frac{1}{2}$ is the probability of getting a head, which is equal to the probability of getting a tail, which is $\frac{1}{2}$.

It is convenient to use **random variables**, which we think of as objects that take a different value, dependent on the outcome of our experiment. For example, in our dice roll we might be interested in the random variable

$$X = \begin{cases} 0 & \text{if the result of the dice roll is less than or equal to 4} \\ 1 & \text{if the result of the dice roll is equal to 5 or 6.} \end{cases}$$

In most cases, it is convenient to let the sample space Ω slip into the background and to just define which random variables we are interested in. So in practice we usually work with random variables, which take random values, which we write with capital letters X, Y etc., and $\mathbb{P}$ which tells us how likely our random variables are to take their various possible values.

1.2 Independence

Sometimes random variables affect each other. For example, in our dice roll, suppose that we defined a random variable Y to be equal to 1 if the dice was a 6, and equal to 0 otherwise. Then, if we know the value of Y we can take a good guess at the value of X.

Sometimes, however, random variables don't affect each other. If we roll the same (fair) dice twice, we don't expect the first roll and the second roll to affect each other. To express this distinction mathematically we use **independence**. The random variables X and Y are independent if

$$\mathbb{P}[X = x, Y = y] = \mathbb{P}[X = x]\mathbb{P}[Y = y]$$

for all values x, y taken by X and Y. This equation says that knowing the value of X doesn't tell us anything about the value of Y, and vice versa.

1.3 Examples of random variables

There are some random variables that are especially useful. These tend to appear in many different situations and have names. We recall some of these special random variables here.

Often random variables, like in our example of a dice, take values from within the integers $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, 3, \ldots\}$ or the natural numbers $\mathbb{N} = \{1, 2, 3, \ldots\}$. We call these random variables **discrete random variables**. Here is one important example (there are many others!):

- **Geometric** random variables, which have a parameter $p \in [0, 1]$. They take random values in $\mathbb{N}$.

  We say that X is a Geom(p) random variable if, for all $n \in \mathbb{N}$,

  $$\mathbb{P}[X = n] = (1 - p)^{n-1} p.$$

  The idea is that we are carrying out an experiment that either succeeds (with probability p) or fails (with probability $1 - p$. A geometric random variable counts the number of times we have to repeat an experiment until we get a success.

If we want to calculate probabilities associated to discrete random variables, we can just add up the probabilities of taking individual values. For example, if X is a Geom(p) random variable,

$$\mathbb{P}[X \leq 2] = \mathbb{P}[X = 1] + \mathbb{P}[X = 2] = p + (1 - p)p = 2p - p^2.$$

It's important to remember that random variables don't *have* to be random. For example, the 'random' variable X might always take the value 0 i.e. $\mathbb{P}[X = 0] = 1$. These are called **deterministic** random variables.

Another type of random variables are ones that take values across all (or part) of the real numbers $\mathbb{R}$. These are known as **continuous** random variables. A continuous random variable Y is usually defined using a **probability density function** (or **p.d.f.**) $f_Y : \mathbb{R} \to \mathbb{R}$, and to calculate probabilities with Y we would use the formula

$$\mathbb{P}[y_1 \leq Y \leq y_2] = \int_{y_1}^{y_2} f_Y(y)\, dy.$$

We can't just 'add up the values' because there are uncountably many values in the interval $[y_1, y_2]$. Instead, we integrate.

Let's look at two important examples of continuous random variables:

- Exponential random variables, which have a parameter $\lambda \in (0, \infty)$. We say that Y is an Exp(λ) random variable if it has the p.d.f.

  $$f_Y(y) = \begin{cases} \lambda e^{-\lambda y} & \text{if } y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

- Normal random variables, which have two parameters $\mu \in \mathbb{R}$ and $\sigma \in (0, \infty)$. We say that Y is a $N(\mu, \sigma^2)$ variable if it has the p.d.f.

  $$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y - \mu)^2}{2\sigma^2}\right).$$

So, calculating probability for continuous random variables comes down to calculating integrals. In some cases (such as normal random variables) this can't always be done 'by hand' and numerical approximations have to be used.

If is common to use the symbol $\sim$ to define a random variable, for example we might write $Y \sim N(\mu, \sigma^2)$ or $X \sim \text{Geom}(p)$.

2 Expectation and variance

Random variables can have very complicated ranges of values and probabilities, so it's useful to be able to summarize them. The two most important features of a random variable are its **mean** and **variance**.

We need to define mean and variance separately for our discrete and continuous case. In the discrete case, the mean is

$$\mathbb{E}[X] = \sum_x x\mathbb{P}[X = x] \tag{2.1}$$

where the $\sum$ is over the values that X takes. As an exercise, you might like to check (or remind yourself from your old notes) that the mean of a Geom(p) random variable, which from (2.1) is equal to $\sum_{n=1}^{\infty} n(1-p)^{n-1}p$, is $\frac{1}{p}$. The variance is

$$\text{var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right], \tag{2.2}$$

which makes sense because $Z = (X - \mathbb{E}[X])^2$ is itself a random variable, so we can use (2.1) to calculate $\text{var}(X) = \mathbb{E}[Z]$. The interpretation of mean and variance is that the value of X are centered around the mean value $\mathbb{E}[X]$, and $\text{var}(X)$ measure how close the values of X tend to be to $\mathbb{E}[X]$.

You will hopefully remember (if not, check it in your old notes!) that

$$\text{var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]. \tag{2.3}$$

In the continuous case, if Y has p.d.f. $f_Y(y)$ then by definition

$$\mathbb{E}[Y] = \int_{-\infty}^{\infty} y f_Y(y)\, dy. \tag{2.4}$$

The variance is again defined using the formula (2.2) (with Y in place of X) and the formula (2.3) holds in this case too.

Again, you might like to check or remind yourself that if $Y \sim \text{Exp}(\lambda)$ then $\mathbb{E}[Y] = \text{var}(Y) = \frac{1}{\lambda}$ and that if $Y \sim N(\mu, \sigma^2)$ then $\mathbb{E}[Y] = \mu$ and $\text{var}(Y) = \sigma^2$.

2.1 Random variables without a mean

Calculating $\mathbb{E}[X]$ using (2.1) and (2.4) asks us to evaluate a sum or an integral. This is not always possible; series do not always converge and integrals do not always exist. This causes the (at first glance, surprising) existence of random variables that do not have a mean value, or whose mean value is equal to $\pm\infty$.

A simple example is the continuous random variable with p.d.f.

$$f_Y(y) = \begin{cases} \frac{1}{y^2} & \text{if } y > 1, \\ 0 & \text{otherwise.} \end{cases}$$

Then, $\mathbb{E}[Y] = \int_1^{\infty} y y^{-2}\, dy = \int_1^{\infty} y^{-1}\, dy = [\log y]_1^{\infty} = \infty$. We would say that Y has **infinite mean**.

Similarly, some random random have finite variance and others do not. For a fixed $k \in \mathbb{N}$, some have finite $\mathbb{E}[X^k]$ and others do not. Generally, random variables with finite expectation (and variance) are easier to work with than those which don't. But in many situations we can't avoid using random variables with infinite mean and/or variance.

2.2 Conditioning

An important technique is conditioning random variables on whether or not particular events happen. One way to do this is by using **indicator functions**, which may be new to you. If E is some event then the indicator function of E is the random variable

$$\mathbb{1}_E = \begin{cases} 1 & \text{if } E \text{ occurs} \\ 0 & \text{if } E \text{ does not occur.} \end{cases}$$

Note that this means that $\mathbb{E}[\mathbb{1}_E] = \mathbb{P}[E]$.

If we have some random variable X, we can then write

$$X = \mathbb{1}_E X + \mathbb{1}_{E^c} X.$$

The term $\mathbb{1}_E X$ is equal to X is E happens, and otherwise zero. Similarly the term $\mathbb{1}_{E^c} X$ is equal to X is E doesn't happen, and otherwise zero.

This technique is useful when we want a particular event to happen. An example of this is Markov's inequality:

Lemma 2.1 *Let X be a random variable and suppose $X \geq 0$. Let $a > 0$. Then*

$$\mathbb{P}[X \geq a] \leq \frac{1}{a}\mathbb{E}[X]. \tag{2.5}$$

The idea here is that a is (typically) large and we want to control the chance of X being larger than a. So we want the 'bad' event $E = \{X \geq a\}$ to happen. We condition: $X = \mathbb{1}_{X<a} X + \mathbb{1}_{X \geq a} X$ and then

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}[\mathbb{1}_{X<a} X + \mathbb{1}_{X \geq a} X] \\ &= \mathbb{E}\left[\mathbb{1}_{X<a} X\right] + \mathbb{E}\left[\mathbb{1}_{X \geq a} X\right] \\ &\geq \mathbb{E}\left[\mathbb{1}_{X \geq a} X\right] \\ &\geq \mathbb{E}\left[\mathbb{1}_{X \geq a} a\right] \\ &= a\mathbb{P}[X \geq a]. \end{aligned}$$

Here we use that $\mathbb{E}\left[\mathbb{1}_{X<a} X\right] \geq 0$, and that if $\mathbb{1}_{X \geq a} X$ is not equal to zero then $X \geq a$. Dividing through by a gives (2.5).

3 Analysis

3.1 sup and inf

You should be familiar with using $\inf A$ (the greatest lower bound of A) and $\sup A$ (the lowest upper bound of A).

Intuitively, you can think of sup as a version of max that doesn't fail to work if you give it something like $[0, 1)$ – it can see that $[0, 1)$ is open ended at the top and so has no 'proper' maximum, but that 1 is the 'next best thing' to having a maximum value. You can think of inf in the same way, as an improved version of min that could handle e.g. $(-2, 1]$ and give $\inf(-2, 1] = -2$.

3.2 Convergence

You should be familiar with the formal 'ϵ, N' definition of convergence for real numbers. The intuition is that if a_n and a are real numbers, then $a_n \to a$ means: as n increases, the value of a_n becomes close to the value of a.

A new idea that you might like to think about is how to apply this idea to random variables. Real numbers are very simple objects – they are a single value. A random variable is a more complicated object since it takes multiple values with different probabilities. But it makes sense intuitively that two random variables can be approximately equal, so it 'should' be possible to think of a sensible meaning for $X_n \to X$.

We'll study it in the course, but let us give a quick example to show that convergence of random variables need a bit of thought.

We might hope that (1) if 'X_n converged to X' that, for any a that is small and positive, the quantity

$$\mathbb{P}[|X_n - X| > a]$$

tended to zero as $n \to \infty$. Equally, we might hope that (2) $\mathbb{E}[X_n] \to \mathbb{E}[X]$.

Let X_n be a random variable such that $\mathbb{P}[X_n = n^2] = \frac{1}{n}$ and $\mathbb{P}[X_n = 0] = 1 - \frac{1}{n}$. Now, if we define also the deterministic random variable $X = 0$, using (2.5) we have

$$\mathbb{P}[|X_n - X| > a] = \mathbb{P}[X_n = n] = \frac{1}{n}$$

which does become small as $n \to \infty$, because $\frac{1}{n} \to 0$.

However, $\mathbb{E}[X_n] = n^2 \cdot \frac{1}{n} = n \to \infty$ as $n \to \infty$, but $\mathbb{E}[X] = 0$. So, in this case (1) happened and (2) didn't. With a little work (you might like to try it!) its possible to think of a sequence Y_n of random variables such that (2) works and (1) doesn't (with $X = 0$).

We'll look at how to define $X_n \to X$ as part of the course, but the moral for now is: convergence of random variables needs more thought that convergence of real numbers.